

ABSTRACT OF THE DISCLOSURE

Approximate answers to queries are provided by executing queries against a representation of a data source in addition to, or in lieu of accessing the source data itself. A representation of a data source, referred to herein as an Information Reservoir, is constructed and maintained using probabilistic methodologies based upon a Poisson sampling approach. The Information Reservoir provides approximate answers to ad hoc queries, potentially in a small fraction of the time required to calculate an exact answer. Associated variances are also provided that may additionally be used to calculate confidence intervals bounding the exact answer. An Information Reservoir may be biased toward a subset of the information in the original data source and/or tailored to the anticipated query workload. Queries expressed as if directed to the original data source may be automatically translated to run against the Information Reservoir with little or no additional burden placed on the Information Reservoir user. Information Reservoir collections may be created that offer users approximate answers of varying levels of precision. Information Reservoirs may also be combined with non-sampling concise representations to increase the precision of approximate answers for certain classes of queries. For example, approximations to specific multidimensional histograms may be combined with an Information Reservoir to accommodate highly selective queries that sampling does not effectively address.

20